
The Role of Databases in Support of Computational Chemistry Calculations

DAVID FELLER*

Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, Washington 99352

Received 3 July 1995; accepted 29 November 1995

ABSTRACT

A role for electronic structure databases in assisting users of quantum chemistry applications select better model parameters is discussed in light of experiences gained from a software prototype known as the Computational Chemistry Input Assistant (CCIA). It is argued that the ready availability of information pertaining to the applications and theoretical models can substantially increase the likelihood of novice users obtaining the desired accuracy from their calculations while simultaneously making better use of computer resources. Expert users, who find themselves contemplating studies in new areas of research, may also benefit from the proposed tools. For maximum impact, this assistance should be provided while users are actively engaged in preparing calculations. © 1996 by John Wiley & Sons, Inc.

Introduction

Sustained rapid advances in microprocessor technology and the widespread availability of sophisticated electronic structure programs have combined to dramatically increase the number of scientists performing theoretical calculations in support of their research activities. Although the majority of these scientists have limited formal training in theoretical chemistry, the ease of use of *ab initio* programs has improved to the point where calculations which were initially the exclusive domain of professional theoreticians are now acces-

sible to even novice users. Greater numbers of *ab initio* programs are being supported by commercial graphical user interfaces, which replace the tedious, error-prone editing of input files with simple "point and click" operations. High resolution color displays and fast graphics hardware, available on systems ranging from low-cost, personal computers to more expensive UNIX workstations, have been harnessed for a variety of molecular visualization tasks. Prominent examples include a wide variety of "builders" to visually specify a chemical system via predefined molecular fragment libraries and three-dimensional displays of electron densities, electrostatic potentials, molecular dynamics simulations, and many others. Nevertheless, while skillfully applied visualization

* Author to whom all correspondence should be addressed.

techniques can help transform large volumes of difficult-to-comprehend, raw, numerical output into more easily assimilated pictures, they do little to address most of the underlying complexities inherent to this type of molecular modeling.

Electronic structure methods, especially of the *ab initio* variety (and we include here recently developed density functional methods), can be enormously powerful probes of molecular phenomena, but the quality of the computed properties is critically dependent on the quality of the instructions supplied to the application software. All too often, for reasons of computational expediency and the lack of practical *a priori* error estimates, this dependence is glossed over. As an undesirable consequence, users may find themselves in a situation where they have provided a syntactically correct input file for performing the desired calculation, the application may have obtained a valid solution to the proper equations and, yet, the output from the calculation may be essentially meaningless due to a failure to understand how the results depended on the choice of input parameters. Because the calculation violated no identifiable rule, which might have triggered a warning message in the output file, users (especially novice users) may proceed to draw incorrect conclusions based on the assumed accuracy of their findings.

Each new release of major program packages such as Gaussian,¹ GAMESS,² or MOLPRO³ normally includes a collection of new features and the associated input parameters to allow users to access the new features. Because these programs have development histories stretching over several decades, the accumulated list of features is quite long and the total number of input parameters reaches into the hundreds. Fortunately, if we ignore the data needed to specify the molecular geometry, the number of *primary* parameters which determine the quality of the calculated properties is far less. For the majority of calculations the primary parameters are: (1) the basis set of functions required for expanding the molecular orbitals or electron density; and (2) the "level-of-theory" or degree of sophistication of the underlying mathematical model. Proper choices for both are critical if one wishes to achieve acceptable accuracy within the constraints imposed by finite computer resources. After the primary input parameters have been selected, the scientist must choose an application and hardware platform on which to run the calculation.

In the U.S. Department of Energy's new Environmental Molecular Sciences Laboratory (EMSL), there are currently more than a dozen *ab initio* codes in use and several hundred workstations, including models from all of the major hardware vendors. Furthermore, for production calculations scientists have access to traditional supercomputer facilities at the National Energy Research Supercomputer Center and, in the future, parallel processing machines. The computational choices confronting the user are complex.

The present work focuses on the role systematically defined databases can play in assisting users in the selection of appropriate basis sets, levels of theory, and hardware/software combinations. It also discusses how information related to the characteristics of specific applications can ease transitions among programs, during the course of a research project, in cases where no single package offers all of the required functionality. For the sake of the present discussion, the term "database" will be used in the generic sense to mean any "large collection of data in a computer, organized so that it can be expanded, updated, and retrieved rapidly for various uses."⁴

To explore the feasibility of the concepts described below, a "proof-of-concept" prototype was developed around a collection of simple ASCII-formatted files and associated support utilities. Although much of the essential chemistry functionality was achievable without resorting to traditional database applications, the purpose of the prototype was not to produce commercial quality robustness. The Computational Chemistry Environment (CCE) Group at the Pacific Northwest Laboratory is developing a suite of UNIX workstation software incorporating a sophisticated chemistry data model and commercial, object-oriented database software. Whenever appropriate, areas of common functionality between the Extensible Computational Chemistry Environment (ECCE) software and the prototype, where the former extends the capabilities of the latter, will be noted.

Previous work in this area includes that of Cushing and coworkers⁵⁻⁷ who discussed the concept of "computational proxies" within the context of a UNIX prototype implemented in C++ using a commercial object-oriented database. Their work focused on the role of database software systems in assisting the research efforts of computational chemists.

Although the present discussion focuses on the field of *ab initio* quantum chemistry, many semi-

empirical methods, which explicitly avoid the complications arising from basis set expansions, can be represented under the same hierarchical input parameter model. Among the scores of semiempirical user-alterable input parameters are a few (e.g., the choice of model Hamiltonian) which will primarily determine the accuracy of the calculation. Fortunately, as with *ab initio* methods, the majority of the less important parameters can be assigned default values by the program authors, with no adverse effect on the calculation.

Thus, the majority of electronic structure models of molecular phenomena present the user with a relatively small number of primary parameters, and a much larger set of secondary parameters whose values are less critical to the quality of the results. By providing interactive assistance in selecting suitable primary parameters at the input preparation stage of a calculation, the chances of obtaining meaningful results are enhanced. In the present work we describe the design of a suite of software utilities which constitute a high-level help facility for users of *ab initio* electronic structure methods.

Basis Sets

Although a small number of quantum-mechanical methods, such as those designed specifically for diatomic molecules, can avoid the need for basis set expansions, the overwhelming majority utilize Gaussian functions $g(x, y, z) = x^l y^m z^n \exp(-\zeta r^2)$ as a mathematical expediency. Newly developed density functional theory (DFT) methods and resolution of the identity perturbation theory (RI-MP2),^{8,9} may even employ multiple basis sets in a single calculation. Despite their ubiquitous nature and their importance in determining the overall quality of a calculation, basis sets are often mentally relegated to the same catch-all class of input parameters as the other options. The tendency is to consider them a technical detail.

The chemistry literature contains references to hundreds of Gaussian basis sets, and more are being added every year. The driving force behind the continued development of new sets is the desire to achieve improved results with less computer time. Since the time to perform correlated *ab initio* calculations scales with the fifth power of the number of basis functions (or worse), even small reductions in the size of a basis set can dramatically affect the time to solution. Another factor is the wide variety of properties that can be

computed. Basis sets which perform well for one property may do very poorly for another. Some properties require more near-valence diffuse functions, while others require higher angular momentum functions. Thus, basis set development is likely to continue to be a fertile area of research.

While, in general, larger basis sets provide better results, the situation is anything but simple. All too often a "bigger" basis set can actually lead to poorer agreement with experiment. By carefully constructing a sequence of ever-larger basis sets which spans all of space, one can approach the complete basis set (CBS) limit. In theory this would require an infinite number of functions, but in practice the limit can be approximated reasonably well with very extended sets. The CBS limit represents the exact solution for the chosen level of theory and, as such, is the accuracy goal which computational chemists strive to attain.

To help users deal with the plethora of basis sets, most *ab initio* packages provide embedded basis set libraries that eliminate the chore of manually entering lists of exponents and contraction coefficients. Typically, these libraries support a selected subset of the most popular basis sets. In some cases, the basis set library is contained in one or more ASCII files, external to the program. Otherwise it is stored within data structures internal to the compiled application. In the latter case, if the user desires a basis set which was not provided by the application's authors, the only alternative is to enter the basis set along with the other input parameters. In either case, if users need new, or less frequently chosen basis sets, they are forced to learn yet another application-unique syntax.

By creating an application-independent Gaussian Basis Set Database, augmented with information about the basis sets and the supported applications, many of the more arcane basis set tasks can be eliminated. A prototype Gaussian basis set database and accompanying graphical user interface were implemented as a vehicle for testing and refining the desired functionality with volunteer scientists from the Environmental Molecular Sciences Laboratory. Although most of the testing was done on an informal basis, a formal usability test was conducted several years into the project. One of the earliest versions of the basis set tool utilized a commercial relational database. However, this approach was quickly abandoned in favor of writing our own database query software built around a set of simple ASCII files.

The total size of the database is quite small by commercial standards, amounting to less than 5

MB. Nevertheless, access speed on slower personal computers was a design issue since the prototype interface, known as the Computational Chemistry Input Assistant (CCIA), was implemented in Allegiant SuperCard^{®10} on an Apple Macintosh[®]. Therefore, it was decided that each family of basis sets (e.g., STO-3G, 3-21G, cc-pVDZ, etc.) should be stored in a separate file to improve access times and to facilitate data management. The ability to perform global changes across all basis sets, which would have argued for a single large file, was considered a very low priority item. Furthermore, a single large and growing file can become awkward to handle. Beyond the division of basis sets by family, additional improvements in access time were achieved by creating a master table file that contains: (1) the list of elements each basis set supports; (2) the maximum *l* value for each set; (3) an intended use label (orbital, DFT fitting, polarization, etc.); (4) associated effective core potentials (ECPs); and (5) miscellaneous reference data. For most database inquiries the individual basis set files are not even accessed, since the master file contains the relevant information.

As of this writing the database supports 106 basis sets, including several DFT orbital and fitting sets, and 3 families of ECPs (see Table I for examples) and 12 different applications, such as GAMESS,² GAMESS(UK), Gaussian-92,¹ MOLPRO,³ HONDO,¹¹ MELDF,¹² SUPERMOLECULE,¹³ NWChem,¹⁴ MOLCAS,¹⁵ ACESII,¹⁶ TurboMole,¹⁷⁻¹⁹ and TX93.²⁰ The rate of introduction of

new sets is still running at 10 to 20 sets per year. Areas such as DFT charge and exchange correlation fitting sets and periodic Hartree-Fock basis sets offer considerable room for growth.

In Figure 1 the opening window of the basis set tool interface is shown as it would appear if the user had specified a hydrocarbon for the chemical system. In simple cases, e.g., when a single basis set suffices for all elements in a molecule, the only required action on the part of the user is to double click the desired basis set from the scrolling list at the left of the window. This level of functionality is already found in most commercial products. However, the additional database-resident information pertaining to basis sets and applications permits the user interface to perform a wider variety of activities, many of which would otherwise be impossible. Representative examples include:

- Graphically displaying which elements are supported by each basis set (each supported element has a small colored triangle in the lower left hand corner).
- Ensuring that the selected basis set does not exceed the maximum number of Gaussian primitives per contracted basis function or the maximum *L* value for the chosen application. If necessary, the offending functions can be removed or, in some cases, transformed so that they do not exceed the application's limit.

TABLE I.
Selected Basis Sets Contained in the Database of the Gaussian Basis Set.^a

STO-2G	Huzinaga MINI	cc-pCVDZ
STO-3G	MINI (scaled)	cc-pCVTZ
STO-6G	Huzinaga MIDI	cc-pCVQZ
3-21G	Dunning / Hay SV	cc-pCV5Z
3-21++ G	Dunning / Hay SVP	NASA Ames ANO
3-21G*	Dunning DZ	Chipman DZP
3-21++ G*	Dunning DZP	Roos Aug. DZ ANO
4-31G	cc-pVDZ	Roos Aug. TZ ANO
6-31G	cc-pVTZ	Ahlrichs pVDZ
6-31G*	cc-pVQZ	Binning / Curtiss SV
6-31G**	cc-pV5Z	Wachter's 1st Row Transit. Metals + f's
6-31++ G**	cc-pV6Z	Bauschlicher 1st Row Transit. Metals
6-311G**	aug-cc-pVDZ	Hay-Wadt MB ECP
6-311++ G(2d, 2p)	aug-cc-pVTZ	Hay-Wadt VDZ ECP
6-311G(2df, 2pd)	aug-cc-pVQZ	SBKJC VDZ ECP
6-311++ G(3df, 3pd)	aug-cc-pV5Z	DeMon DFT Charge Fitting Basis

^a The total number of basis set families contained in the database as of 11 / 16 / 95 is 106.

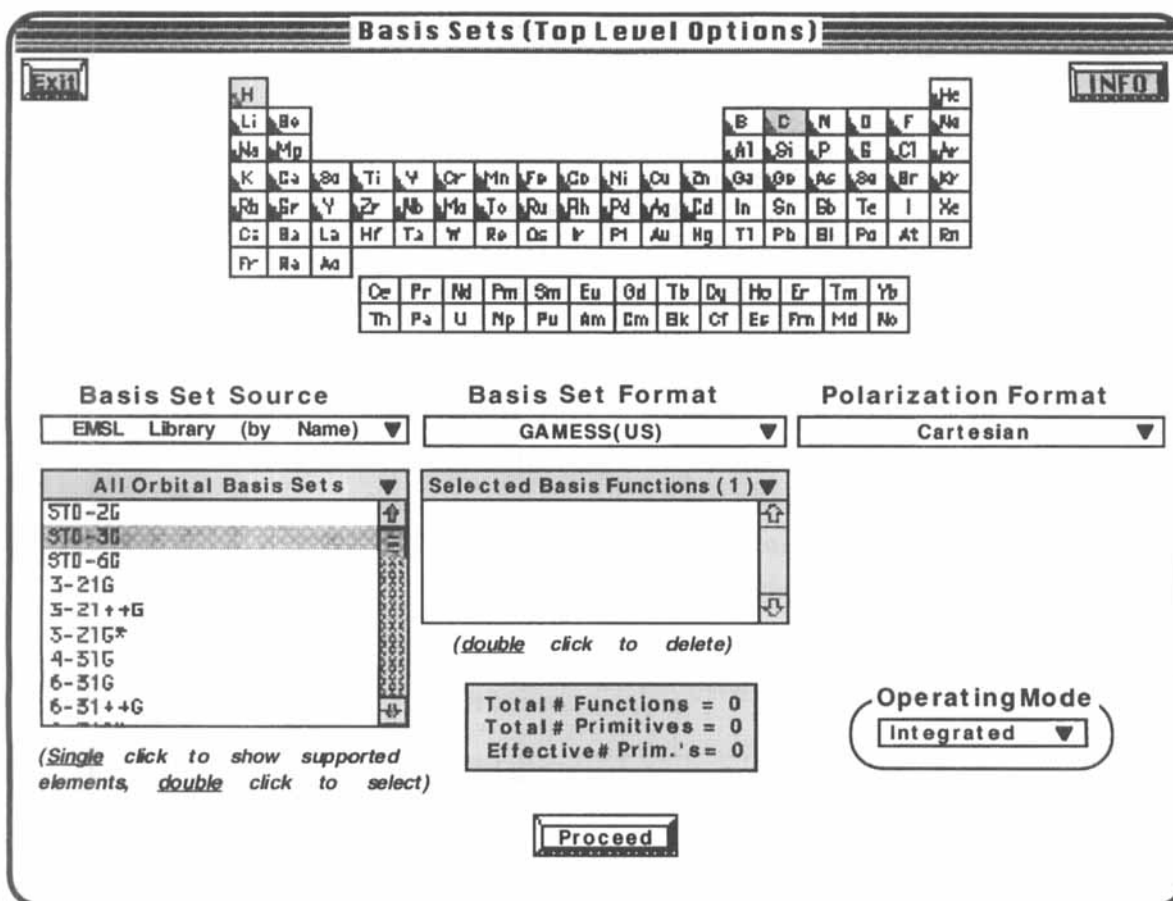


FIGURE 1. The CCIA prototype Gaussian Basis Set Utility top level window.

- Tracking which basis sets were designed for spherical vs. Cartesian polarization components and which applications support spherical components and informing the user of this feature.
- Ensuring that the selected basis set supports all of the elements present in the molecule, or (optionally) displaying only those basis sets which simultaneously support all elements in a given chemical system.
- Tracking which applications do not support the generally contracted basis set and providing a transformation of the general contractions that improve run times.
- Allowing users to mix orbital basis sets (e.g., Dunning/Hay SV) with polarization or diffuse sets taken from another basis.
- Allowing users to select different basis sets for different elements in the same chemical system.
- Providing a range of "help" and associated information pertaining to Gaussian basis sets in general or to a particular basis set (see Fig. 2).

In some cases, differences will arise among programs regarding the definition of certain basis sets. For example, the 6-31 G Si basis set found in GAMESS does not match the 6-31G Si basis contained in Gaussian. A centralized database can help arbitrate differences in naming conventions by enforcing an application-neutral set of names. The CCIA basis set utility provides the user with the option of requesting a basis set from the application's internal library (via the Basis Set Source pulldown menu in Fig. 1) so that none of the user's current capabilities are lost.

The database support utilities, which were implemented in Fortran-77, are portable across most UNIX, PC, and Macintosh platforms. For example, the same CCIA utilities running on the Macintosh

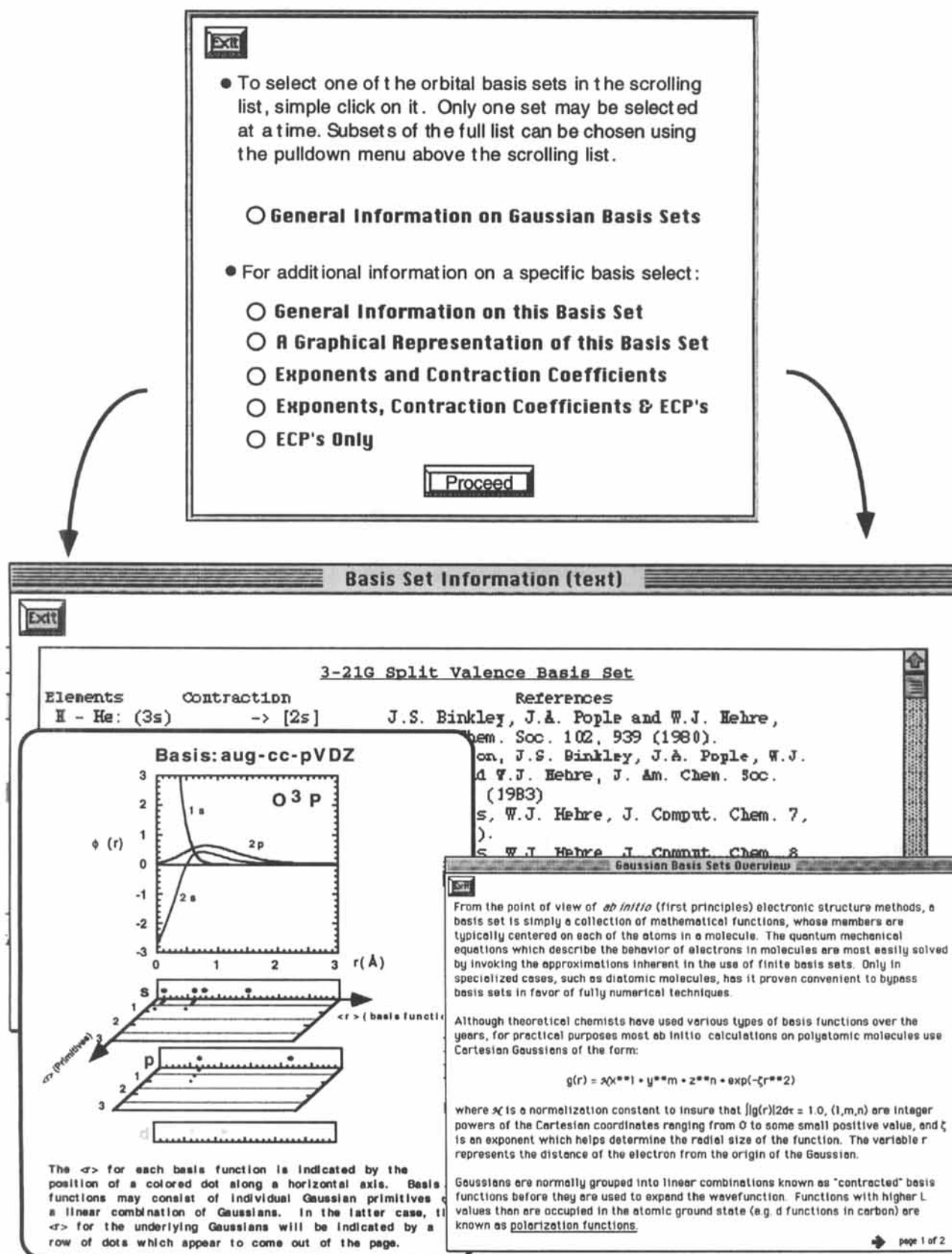


FIGURE 2. Examples of information windows that can be requested from the Gaussian Basis Set Utility.

are currently being used in a World Wide Web basis set tool (URL <http://www.emsl.pnl.gov:2080>) running on a Sun workstation. Since the actual basis set files are in ASCII format, the only prerequisite for using this software is the availability of a Fortran run time environment. The format of the master index file and all of the basis set files (see Table II) were designed to be self-documenting to the extent that they could be easily expanded by individual scientists using nothing more than a simple ASCII text editor. Thus, the scientist is not required to become knowledgeable about a potentially complex database application.

Levels of Theory

The second primary input parameter is informally referred to as the "level of theory" or the

"level of correlation recovery" and controls the sophistication with which the electronic degrees of freedom are treated during the calculation. In terms of traditional *ab initio* levels of theory, the low end (i.e., least sophisticated) is represented by the various Hartree-Fock models. This level of theory has been shown to be adequate for qualitatively describing a wide range of chemical phenomena. However, the apparent accuracy provided by small basis set Hartree-Fock calculations frequently relies on a fortuitous cancellation of errors and is inadequate for many important classes of properties (e.g., accurate bond energies, transition moments or spin properties, to name just a few). As a consequence, theoreticians have devoted a lot of time to developing "correlated" methods which relax the single determinant restriction inherent to Hartree-Fock theory and yet remain computationally tractable for larger chemical systems.

TABLE II.
A Sample Entry in the Database of Gaussian Basis Sets.

NAME = "cc-pVDZ" ! Correlation Consistent Polarized Valence Double Zeta Basis

REFERENCES = "T. H. Dunning, Jr., J. Chem. Phys. 90,1007 (1989).

D. E. Woon and Dunning, Jr., J. Chem. Phys. 98, 1358 (1993).

Na, Be: D. E. Woon and T. H. Dunning, Jr. (To Be Published)"

Z = 6 ! CARBON (Certified 8 / 5 / 91 DFF)

DESCRIPTION = "(9s, 4p, 1d) → [3s, 2p, 1d]"

E(ROHF) = -37.682391 **STATE** = "3-P"

E(SD-CI) = -37.759783 **STATE** = "3-P"

! S-Block		Contraction Coefficients		
Exponents		S	S	S
	6665.0	0.000692	-0.000146	0.0
	1000.0	0.005329	-0.001154	0.0
	228.00	0.027077	-0.005725	0.0
	64.710	0.101718	-0.023312	0.0
	21.060	0.274740	-0.063955	0.0
	7.4950	0.448564	-0.149981	0.0
	2.7970	0.285074	-0.127262	0.0
	0.5215	0.015204	0.544529	0.0
	0.1596	-0.003191	0.580496	1.0
! P-Block		Contraction Coefficients		
Exponents		P	P	
	9.439	0.038109	0.0	
	2.002	0.209480	0.0	
	0.5456	0.508557	0.0	
	0.1517	0.468842	1.0	
! D-Block		Contraction Coefficients		
Exponents		5D		
	0.550	1.0		

! COMMENTS

! The atomic set was obtained by optimizing the (s, p)-exponents in Hartree-Fock calculations on the ground, (3)P, state of carbon. The polarization set was obtained by optimizing the exponent of a single d primitive in SD-CI calculations on the (3)P state. The SD-CI calculations were carried out in D(2h) symmetry; the CI configurations were generated relative to all three (3)P configurations.

At the high end (i.e., most sophisticated) is full configuration interaction (FCI), which scales factorially with the size of the system and can, thus, only be used on small molecules. Spanning the gap between Hartree-Fock theory and full CI is an often bewildering assortment of intermediate-level methods whose names read like an alphabet soup of quantum chemistry. Of the many levels of theory proposed over the years, most have never developed a sizable following. Only a dozen or so correlated methods continue to enjoy widespread use. Newer, more accurate and/or efficient techniques can be expected to gradually displace older methods over time. Keeping track of which current methods are best suited for a particular class of properties can easily overwhelm a full-time computational chemist, much less the occasional user of these methods.

While the choice of a basis set and level of theory are often regarded as two independent variables (some textbooks represent them as two orthogonal axes), in practice it is better to view them as closely coupled. This interdependence is emphasized by the labels "1-particle basis set" and "*n*-particle basis set" which are sometimes used in place of "basis set" and "level of theory." Ideally, there should be a balance between the level of theory and basis set quality to make efficient use of computer resources. Examples of unbalanced calculations can be best illustrated through two extreme cases: (1) small basis sets used in combination with correlated methods; or (2) very large basis sets used at the Hartree-Fock level. Both situations are undesirable, since each employs a lopsided basis set/level of theory combination that lacks comparable flexibility in the other primary input parameter.

Users of electronic structure techniques have long suffered from the lack of rigorous error bars which would allow them to anticipate the likely accuracy of a given basis set/level of theory combination. In the absence of practical theoretical means for bracketing the complete basis set, full CI value of a property, computational chemists have often chosen to calibrate basis set/level of theory combinations on limited classes of molecules chosen so as to mimic the systems of interest. The success of such an approach rests on the assumption that a given theoretical treatment should yield similar errors for the calibrating set of "closely related" molecules, as well as for the original molecules of interest.

When unexpected exceptions to this rule of "similar errors in similar chemical systems" arise,

it can generate intense interest on the part of the chemistry community and serve as a periodic reminder of the limitations of generalizing from a limited body of data. However, once the origin of the anomalous behavior is understood and the knowledge is disseminated in the scientific literature, the chemist's understanding of the computational model's strengths and weaknesses grows to encompass the new information. Unfortunately, more and more scientists find themselves without the time or inclination to perform their own calibration studies or to remain current with the burgeoning literature in this field.

In an attempt to address some of these issues, we are developing a Methods Performance Database of *ab initio* results with the goal of providing relevant information to users who are in the process of setting up electronic structure calculations. Currently this database covers 26 of the basis sets described previously and all of the levels of theory listed in Table III. The general areas of focus include: (1) molecular structures; (2) molecular energetics (i.e., dissociation energies, electron affinities, proton affinities, excitation energies, rotation/inversion barriers, and ionization potentials); and (3) molecular vibrational frequencies. A representative sample of the 121 molecules and 76 energetic quantities contained in the database as of 11/16/95 are listed in Tables IV and V.

Using information compiled from the chemistry literature and from thousands of calculations run specifically for this project, more than 6500 entries have been collected in the prototype Methods Performance Database. Occasionally, even, when literature values were available, molecular geometries were reoptimized to assure uniform convergence accuracy in the data. The format of the molecular geometry section of the database is illustrated with a simple example in Table VI, where a collection of water entries is shown. In addition to the body of basis set/level of theory entries, the database section for each molecule also contains the following reference information: (1) a set of experimental values; (2) the best estimate of the complete basis set, Hartree-Fock limit; (3) the best correlated theory values; and (4) literature citations for the reference values. Whenever reference values were unavailable, such as might occur for extended basis set results on very large molecules, or for molecules lacking experimental data, the corresponding entries were zeroed out. To be considered for inclusion in the Methods Performance Database, a molecule or reaction must have either reliable experimental data or high level theoretical results.

TABLE III.
Levels of Theory Represented in the Prototype Basis Set and Methods Performance Database.^a

Method	Description
RHF	Restricted Hartree–Fock
UHF	Unrestricted Hartree–Fock
ROHF	Open-Shell Restricted Hartree–Fock
MP2	Second-Order Møller–Plesset Perturbation Theory
MP4	Fourth-Order Møller–Plesset Perturbation Theory
QCISD	Singles and Doubles Quadratic CI
QCISD(T)	Singles and Doubles Quadratic CI with Perturbative Triples
CCSD	Coupled Cluster with Singles and Doubles
CCSD(T)	Coupled Cluster with Singles and Doubles and Perturbative Triples
CISD	Singles and Doubles CI
CAS-SCF	Complete Active Space Self-Consistent Field
CAS-CI	Complete Active Space Multireference CI
iCAS-CI	Internally Contracted Complete Active Space Multireference CI
FCI	Full CI

^a All correlated methods include both frozen core and nonfrozen core entries.

The vibrational frequency and relative energetics portions of the database follow a similar pattern.

While the number of molecules is relatively small, the emphasis has been on trying to reach 100% coverage for the selected basis sets and levels of theory. Because it is intended to provide the information base underlying the Basis Set/Methods Advisor, the Methods Performance Database must be relatively comprehensive in its coverage. Any significant gap in the database associated with a particular basis set or level of theory will result in a skewing of the recommendations issued from the Advisor. As a consequence, the addition of a new basis set or level of theory entails a significant amount of work, since it must be applied to all (or nearly all) of the molecules and reactions in the database. Many of the larger calculations performed for this project involve larger basis sets and/or higher levels of theory than have previously been published.

Exceptions to this rule can occur for very large basis sets, which would require intractable calculations on large molecules. Nonetheless, a very large number of previously unpublished calculations have been performed with extended basis sets on moderately large systems for the sole purpose of populating the database. We have also been fortunate to be able to exploit the results of Dunning and coworkers^{21–24} who have used basis sets up through quintuple-zeta quality (*g* functions on hydrogen/*h* functions on nonhydrogen atoms). These authors have a growing number of articles

documenting the accuracy of the correlation-consistent family of basis set and the intrinsic accuracy of various levels of theory. The latter was made possible because of the ability to estimate the complete basis set limit using the regularity in properties obtained from the correlation consistent basis sets. Pople and coworkers have established a long-standing tradition in this area.^{25,26} While relying, whenever possible, on previously published results, the Methods Performance Database differs from previous calibration studies in that its scope entails basis sets and levels of theory from many different research groups.

Due to time constraints in the development of the prototype Methods Performance Database, its contents differ in two major ways from the design objective of the ECCE production software. First, the prototype does not provide a complete set of internal coordinates for larger molecules, whereas the production software attempts to establish a pointer to the full list of Cartesian coordinates, if available. Second, the prototype currently does not store citation information for each basis set/level of theory entry. Again, whenever possible the production database will contain information on the author(s) of the entry, as well as miscellaneous information, such as the date and the machine on which the calculation was run, as well as the name of the application. Since this information was judged nonessential for assisting scientists in the choice of input parameters, no attempt was made to include it in the prototype.

TABLE IV.
Selected Molecules Represented in the Prototype Basis Set and Methods Performance Database.^a

H ₂ (¹ Σ _g ⁺)	LiH (¹ Σ ⁺)	BeH (² Σ _g)
BeO (¹ Σ _g ⁺)	BH (¹ Σ ⁺)	BH ₂ (² A ₁)
BH ₃ (¹ A ₁) Planar D3h	B ₂ (¹ Σ ⁺)	BF (¹ Σ ⁺)
BF ₃ (¹ A ₁) Planar D3h	OH (² Π)	HF (¹ Σ ⁺)
NH (³ Σ ⁻)	NH ₂ (² B ₁)	CO (¹ Σ ⁺)
CO ₂ (¹ Σ _g ⁺)	CN (² Σ ⁺)	CP (² Σ ⁺)
O ₃ (¹ A ₁)	Li ₂ (¹ Σ _g ⁺)	LiF (¹ Σ ⁺)
LiCl (¹ Σ ⁺)	N ₂ (¹ Σ _g ⁺)	F ₂ (¹ Σ _g ⁺)
F ₂ O (¹ A ₁) oxygen difluoride	NH ₃ (¹ A ₁)	NH ₄ ⁺ (¹ A ₁)
H ₂ O (¹ A ₁)	H ₂ O ⁺ (² B ₁)	H ₃ O ⁺ (¹ A ₁)
(H ₂ O) ₂ (¹ A') water dimer	CH (² Π)	CH ₂ (³ B ₁)
CH ₂ (¹ A ₁)	CH ₂ ⁻ (² B ₁)	CH ₃ methyl radical (² A ₂)
CH ₄ (¹ A ₁)	O ₂ (³ Σ ⁻)	C ₁₀ H ₁₀ Fe (¹ A ₁) ferrocene
HCO formyl radical (² A')	HCO ₂ H formic acid (cis)	HCN (¹ Σ ⁺)
HNC (¹ Σ ⁺)	LiOH (¹ Σ ⁺)	B ₂ H ₆
C ₂ (¹ Σ _g ⁺)	C ₂ ⁺ (⁴ Σ _g ⁻)	C ₂ ⁺ (² Π _u) excited state
C ₂ ⁻ (² Σ _g ⁺)	C ₂ H ₂ acetylene	C ₂ H ₄ ethylene
CH ₂ NH	CH ₃ CHO acetaldehyde	CH ₃ CO +
C ₂ H ₆ staggered ethane	C ₂ H ₆ eclipsed ethane	C ₃ H ₆ cyclopropane
C ₃ H ₄ cyclopropene	C ₃ H ₈ propane	C ₃ H ₆ propene
C ₄ H ₆ trans-1,3-butadiene	C ₂ H ₅ NO acetamide	C ₃ H ₇ NO methylacetamide
C ₃ H ₅ allyl (² A ₂)	C ₄ H ₁₀ isobutane	C ₄ H ₈ isobutene
C ₆ H ₆ benzene	CF ₄ (¹ A ₁) carbon tetrafluoride	HO ₂
H ₂ CO (¹ A ₁) formaldehyde	CH ₃ OH methanol	C ₂ H ₅ OH ethanol
CH ₃ NH ₂ methylamine	CH ₃ F	C ₁₂ H ₂₄ O ₆ 18-crown-6
H ₂ S (¹ A ₁)	NO (² Π)	HNO (¹ A')
HNCO (¹ A')	N ₂ O (¹ Σ ⁺) nitrous oxide	NH ₂ OH
H ₂ O ₂ (¹ A) hydrogen peroxide	FOOF	FNO ₂
C ₆ H ₅ OH phenol	C ₂ H ₆ O methyl ether	HCl (¹ Σ ⁺)
PN (¹ Σ ⁺)	P ₂ (¹ Σ _g ⁺)	HSO (² A'')
NaF (¹ Σ ⁺)	NaCl (¹ Σ ⁺)	KCl (¹ Σ ⁺)
SiH ₂ (¹ A ₁)	SiH ₂ (³ B ₁)	SiH ₄ (¹ A ₁)

^a As of 11/16/95, information on 121 molecules was stored in the database.

Users seeking help in the choice of basis sets or a level of theory are presented with the window shown in Figure 3. As was the case with the basis set utility (Fig. 1), the chemical system would have been defined prior to the appearance of this window. By specifying the type of information desired from the calculation (e.g., electron affinities or bond stretch frequencies) along with an approximate accuracy, the user directs the software to scan the database for entries which correspond to the current chemical system and simultaneously match any additional search parameters that might have been specified. For example, if the molecule of

interest was a hydrocarbon and the user indicated an interest in computing bond lengths with an accuracy of ± 4 Å and bond angles with an accuracy of $\pm 7^\circ$, the software would search the database looking for all molecules containing carbon and hydrogen. If sufficient entries were discovered, the two basis sets with the smallest normalized root-mean-square (rms) error measured with respect to the requested reference values, i.e., experiment and/or the best available theoretical values, would be recommended (see Fig. 4A). Otherwise, a message would appear stating that there were insufficient data to make a recommendation.

TABLE V.
Representative Entries in the Prototype Basis Set and Methods Performance Database.

Dissociation energies (De)	
$\text{H}_2 (^1\Sigma_g^+) \rightarrow \text{H} (^2\text{S}) + \text{H} (^2\text{S})$ $\text{Li}_2 (^1\Sigma_g^+) \rightarrow \text{Li} (^2\text{S}) + \text{Li} (^2\text{S})$ $\text{LiCl} (^1\Sigma^+) \rightarrow \text{Li} (^2\text{S}) + \text{Cl} (^2\text{P})$ $\text{BeO} (^1\Sigma^+) \rightarrow \text{Be} (^1\text{S}) + \text{O} (^3\text{P})$ $\text{BH} (^1\Sigma^+) \rightarrow \text{B} (^2\text{P}) + \text{H} (^2\text{S})$ $\text{BF} (^1\Sigma^+) \rightarrow \text{B} (^2\text{P}) + \text{F} (^2\text{P})$ $\text{CH}_4 \rightarrow \text{CH}_3 + \text{H} (^2\text{S})$ $\text{NO} (^2\Pi) \rightarrow \text{N} (^4\text{S}) + \text{O} (^3\text{P})$ $\text{C}_2\text{H}_2 + 3\text{H}_2 \rightarrow 2\text{CH}_4$ $\text{C}_2\text{H}_6 + \text{H}_2 \rightarrow 2\text{CH}_4$ $\text{C}_2\text{H}_4 + 2\text{H}_2 \rightarrow 2\text{CH}_4$ $\text{C}_3\text{H}_4 + 4\text{H}_2 \rightarrow 3\text{CH}_4$ $\text{PN} (^1\Sigma^+) \rightarrow \text{P} (^4\text{S}) + \text{N} (^4\text{S})$	$\text{LiH} (^1\Sigma^+) \rightarrow \text{Li} (^2\text{S}) + \text{H} (^2\text{S})$ $\text{LiF} (^1\Sigma^+) \rightarrow \text{Li} (^2\text{S}) + \text{F} (^2\text{P})$ $\text{BeH} (^1\Sigma^+) \rightarrow \text{Be} (^1\text{S}) + \text{H} (^2\text{S})$ $\text{B}_2 (^3\Sigma_u^-) \rightarrow \text{B} (^2\text{P}) + \text{B} (^2\text{P})$ $\text{BH}_2 (^1\text{A}_1) \rightarrow \text{BH} (^1\Sigma^+) + \text{H} (^2\text{S})$ $\text{CH} (^2\Pi) \rightarrow \text{C} (^3\text{P}) + \text{H} (^2\text{S})$ $\text{NH} (^3\Sigma^-) \rightarrow \text{N} (^4\text{S}) + \text{H} (^2\text{S})$ $\text{F}_2 (^1\Sigma_g^+) \rightarrow \text{F} (^2\text{P}) + \text{F} (^2\text{P})$ $\text{CH}-\text{CH} \rightarrow \text{CH} + \text{CH}$ $\text{CH}_3-\text{CH}_3 \rightarrow \text{CH}_3 + \text{CH}_3$ $\text{CH}_3-\text{NH}_2 \rightarrow \text{CH}_3 + \text{NH}_2$ $\text{N}_2 (^1\Sigma_g^+) \rightarrow \text{N} (^4\text{S}) + \text{N} (^4\text{S})$
Miscellaneous reactions	
$\text{C}_3\text{H}_4 + 4\text{H}_2 \rightarrow 3\text{CH}_4$ cyclopropene $\text{C}_2\text{H}_4 + 2\text{H}_2 \rightarrow 2\text{CH}_4$ $\text{H}_2\text{CO} \rightarrow \text{HCOH}$ 1, 2 hydrogen shift	$\text{CH}_2 (^3\text{B}_1) \rightarrow \text{CH}_2 (^1\text{A}_1)$ $\text{CH}_2 = \text{CH}_2 + 2(\text{CH}_4) \rightarrow 2(\text{C}_2\text{H}_6)$ $\text{CH}_3\text{OH} \rightarrow \text{CH}_3 + \text{OH}$
Rotation / inversion barriers	
NH_2-CN inversion barrier $\text{H}_3\text{C}-\text{CH}_3$ rotational barrier	NH_3 Inversion barrier
Electron affinities (EA)	
$\text{CH}_2 (^3\text{B}_1) + \text{e}^- \rightarrow \text{CH}_2^- (^2\text{B}_1)$ $\text{C} (^3\text{P}) + \text{e}^- \rightarrow \text{C}^- (^4\text{S})$ $\text{F} (^2\text{P}) + \text{e}^- \rightarrow \text{F}^- (^1\text{S})$	$\text{C}_2 (^1\Sigma_g^+) + \text{e}^- \rightarrow \text{C}_2^- (^2\Sigma_g^+)$ $\text{O} (^3\text{P}) + \text{e}^- \rightarrow \text{O}^- (^2\text{P})$ $\text{H} (^2\text{S}) + \text{e}^- \rightarrow \text{H}^- (^1\text{S})$
Ionization potentials (IP)	
$\text{C}_2 \rightarrow \text{C}_2^+$ $\text{H}_2\text{O} (^1\text{A}_1) \rightarrow \text{H}_2\text{O}^+ (^2\text{A}_1) + \text{e}^-$ VIP $\text{C} (^3\text{P}) \rightarrow \text{C}^+ (^2\text{P}) + \text{e}^-$	$\text{H}_2\text{O} (^1\text{A}_1) \rightarrow \text{H}_2\text{O}^+ (^2\text{B}_1) + \text{e}^-$ VIP $\text{H}_2\text{O} (^1\text{A}_1) \rightarrow \text{H}_2\text{O}^+ (^2\text{B}_2) + \text{e}^-$ VIP
Proton affinities (PA)	
$\text{H}_2\text{O} (^1\text{A}_1) + \text{H}^+ \rightarrow \text{H}_3\text{O}^+$ $\text{NH}_3 + \text{H}^+ \rightarrow \text{NH}_4^+$	$\text{N}_2 + \text{H}^+ \rightarrow \text{N}_2\text{H}^+$

If an explanation is requested for the particular recommendations being offered, the user would be shown one or more windows similar to Figure 4B, which contains a listing of the rms deviations for each type of geometry parameter recognized by the database. The explanation text that appears near the top of Figure 4B states that the overall error for the first-choice Dunning DZ basis was 80% of the second choice, but the basis set was nearly twice as costly to use. A similar statistical

breakdown of the errors is available for the second choice. When even greater detail is desired, the user can request to see a molecule-by-molecule report of the errors, as shown in Figure 5, where all molecules considered in making the recommendation are listed. By optionally displaying this information, the interface allows the user the freedom to personally override the Advisor algorithm.

The Basis Set/Methods Advisor assigns an overall accuracy score to each basis set or level of

TABLE VI.
A Representative Molecular Structure Set of Entries in the Basis Set and Methods Performance Database.

Molecule = "H2O 1A1" Charge = 0 Elements = O H


Labels =	R(OH)	A(HOH)	
Parameter_Types =	AH	HAH	
Best_Experiment =	0.9572	104.52	
Best_HF =	0.940	106.3	
Best_Correlated =	0.959	104.2	
Basis = STO-2G	RHF	0.997	98.5
Basis = STO-3G	RHF	0.990	100.0
Basis = STO-6G	RHF	0.986	100.0
Basis = 3-21G	RHF	0.967	107.6
		:	
Basis = 6-311++ G(3df, 3pd)	RHF	0.940	106.3
Basis = 6-311++ G(3df, 3pd)	MP2(FC)	0.958	104.2
Basis = 6-311++ G(3df, 3pd)	MP4(FC)	0.959	104.2
Basis = aug-cc-pVTZ	RHF	0.941	106.3
Basis = aug-cc-pVTZ	MP2(FC)	0.961	104.1
Basis = aug-cc-pVTZ	MP4(FC)	0.963	104.1
Basis = aug-cc-pVTZ	CAS(FG)	0.963	104.6
Basis = aug-cc-pVTZ	QCISD(T)	0.962	104.1
		:	


Best_Expt_Reference = W. S. Benedict, N. Gailar, and E. K. Plyer, J. Chem. Phys., 524, 1139 (1956).

Best_HF_Reference = RHF/aug-cc-pVQZ, D. Feller, E. Glendening, R. A. Kendall, and K. Peterson, J. Chem. Phys. 100, 4981 (1994).

Best_Correl_Reference = MP4(FC)/aug-cc-pVQZ, D. Feller, E. Glendening, R. A. Kendall, and K. Peterson, J. Chem. Phys., 100, 4981 (1994).

Basis Sets (Seek Advice)





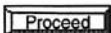
General Topics	Subtopics	Approximate Desired Accuracy	Units
<input checked="" type="checkbox"/> Molecular Structure	All Parameters ▾	±0.04 (dist)/± 7 ▾	Å (1.0E-10m) ▾
<input checked="" type="checkbox"/> Vibrational Modes	All Modes ▾	±500 ▾	cm-1 (4.6E-6Eh) ▾
<input checked="" type="checkbox"/> Relative Energetics	All Energetics ▾	±80.0 ▾	kcal/mol ▾
<input type="checkbox"/> Moments/Polariz.			
<input type="checkbox"/> Transition Moments			
<input type="checkbox"/> Miscellaneous Prop.'s			

Additional Criteria For Making a Recommendation

Types of Sets to Consider <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">All Basis Sets ▾</div>	Consider the Ability to Reproduce <input checked="" type="checkbox"/> Experiment <input checked="" type="checkbox"/> Best Available Theory <input type="checkbox"/> Perform an Application restricted Search	How to Balance Speed vs. Accuracy <input checked="" type="radio"/> Accuracy & Speed <input type="radio"/> Accuracy Alone <input type="checkbox"/> Consider Similarly Charged Species Only
Level of Theory <div style="border: 1px solid black; padding: 2px; margin-bottom: 5px;">Hartree-Fock ▾</div>		

Target Machine

Power Mac 8100/80 ▾






FIGURE 3. The CCIA prototype Basis Set Advisor user interface.

A

Basis Set Recommendations

Exit INFO

First Choice

Name: DZ (Dunning) ☒ Choose Recommended Set 1

Basis Type: DZ Double Zeta: 2 Functions/AO

Second Choice

Name: 3-21++G ☐ Choose Recommended Set 2

Basis Type: VDZD Valence Double Zeta + Diffuse Functions on All

Explanation Proceed Other Sets...

B

Explain Recommendation

Exit

First Choice Basis Set: DZ (Dunning)

The first choice has an overall error which was 80.0% of the second choice but it was 1.8 times slower.

Primitives: 28 # Basis Funct's: 12 Est. CPU min: 0.307 # of Molecules = 44 Show

Recommendation Based on: Hartree-Fock Theory

RMS Geometry Errors (Å and °)

	AH Bonds	AB Bonds	HAH Angles	HAB Angles	ABC Angles	Dihedrals
Number of Comparisons	20	11	9	5	4	0
Compared with Expt.	0.010	0.015	2.389	0.469	7.979	0.000
Compared with HF Limit	0.008	0.032	2.317	0.714	1.712	0.000
Max. Observed Error	0.018	0.026	4.500	0.700	12.200	0.000

Page 1 of (up to) 6 Next Page Prev. Page Go To Page

FIGURE 4. (A) The Recommendation window containing (up to) two basis sets and their general categories from the Basis Set Advisor. (B) One of several explanation windows that the user may request.

theory matching the user's preferences. Because of the need to simultaneously handle dissimilar quantities, e.g., bond lengths, bond angles, and dissociation energies, the rms deviation for each property is approximately normalized to unity by dividing the entries associated with each basis/level of theory combination by the experimental value of the property. If that is unavailable, the best correlated value is used instead. When the user requests the Advisor to consider relative computational costs in addition to agreement with experiment, the three or four candidates with the lowest normalized rms deviations are then subjected to a second sorting on the basis of a combination of rms deviation and a crude estimate of the cost. Currently this estimate assumes a simple N^4 dependence on the number of Gaussian primitives for Hartree-Fock calculations or N^5 for MP2.

Because computation times for all basis sets and methods are estimated with the same degree of

sophistication, the simplicity of the current algorithm was not considered a major drawback. In general, the complexities of integral prescreening and the inherent variations in efficiencies among software packages make the prediction of run times with any accuracy a difficult task. Work is presently under way on the development of a more sophisticated algorithm for predicting the cost of *ab initio* calculations. It will factor in the type of calculation to be performed, the model of computer to be used, and the particular application selected by the user. Information from an extensive benchmarking effort^{27,28} that has been functioning for several years within the EMSL will be stored in a hardware/software performance database. Data has already been tabulated on 11 different program packages and 29 hardware platforms, ranging from high-end personal computers to multimillion-dollar supercomputers and parallel machines. The standard suite of benchmarks ranges

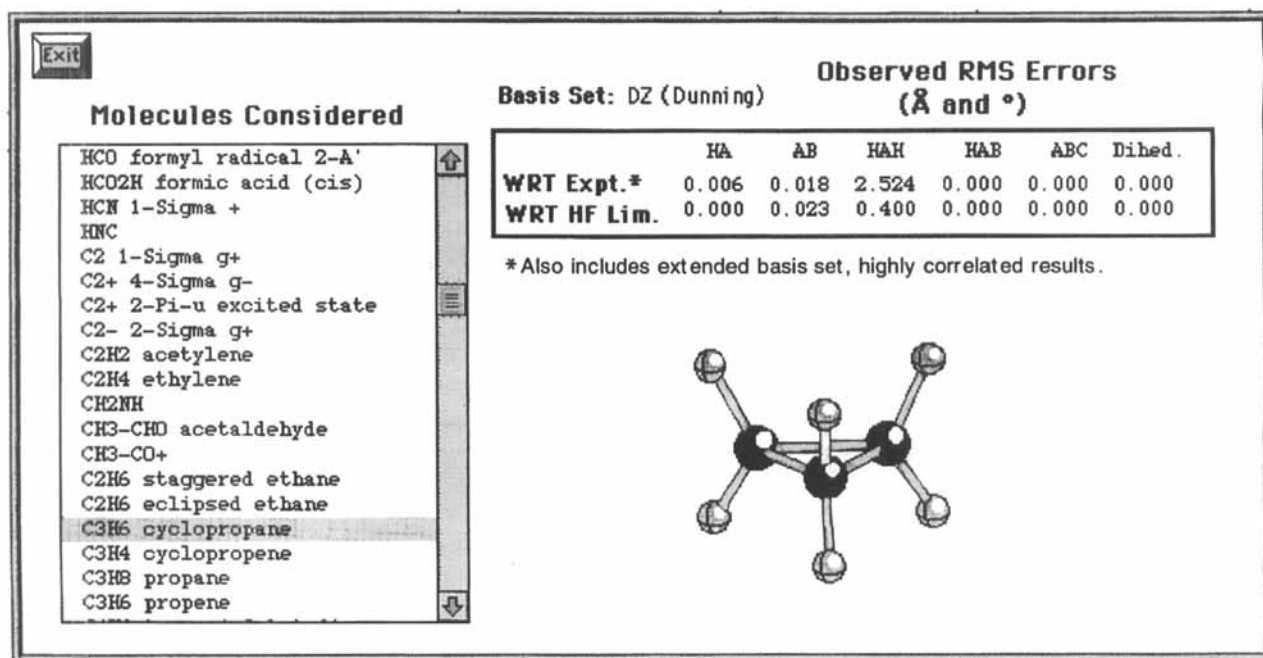


FIGURE 5. A detailed explanation window which lists the observed error with respect to experiment and the Hartree-Fock limit for the Dunning DZ basis at the HF level of theory.

from small molecules with high symmetry to relatively large systems with no exploitable symmetry. It consists of approximately 20 different commonly used levels of theory, including a mixture of direct and conventional algorithms.

More sophisticated users may wish to bypass the Basis Set/Methods Advisor and directly interrogate the Methods Performance Database, since ultimately no piece of software can hope to duplicate the experience gained from years of study. Generic information may be retrieved from the database through the interface shown in Figure 6. The user first selects a category, e.g., "Relative Energetics," and is presented with a list of all entries in the database associated with that category. A pop-up menu allows the user to optionally select a particular topic under the category, e.g., "Dissociation Energies." Other options control the choice of basis set(s), level(s) of theory and the types of information to be obtained. Examples of the types of questions which can be answered by the interface in Figure 6 are:

- For a given basis set, what level of theory is necessary to obtain ± 10 kcal/mol accuracy in the dissociation energy of C=C bonds?
- For a fixed collection of basis sets, does CAS-CI predict better bond lengths than CCSD(T)?

- What singly polarized basis set predicts the best vibrational frequencies for hydrocarbons at the MP2(FC) level?
- What are the RHF/3-21 G C—C bond lengths in isobutane?

By combining this type of information with the previously described benchmark data on the computational cost of various basis set/level of theory combinations, the user should be in a position to make an informed choice of which parameters to select for the problem at hand.

Conclusion

Computational chemistry practitioners at every experience level are increasingly called upon to extend their research activities into new (and therefore unfamiliar) areas of molecular science, due to increased experimental collaborations and the current emphasis on applied research. In the foreseeable future this trend is likely to persist as: (1) computer performance continues to double every 12 to 18 months; (2) additional chemistry software packages enter the marketplace; and (3) the scope of electronic structure projects becomes increasingly ambitious. We have described a suite of

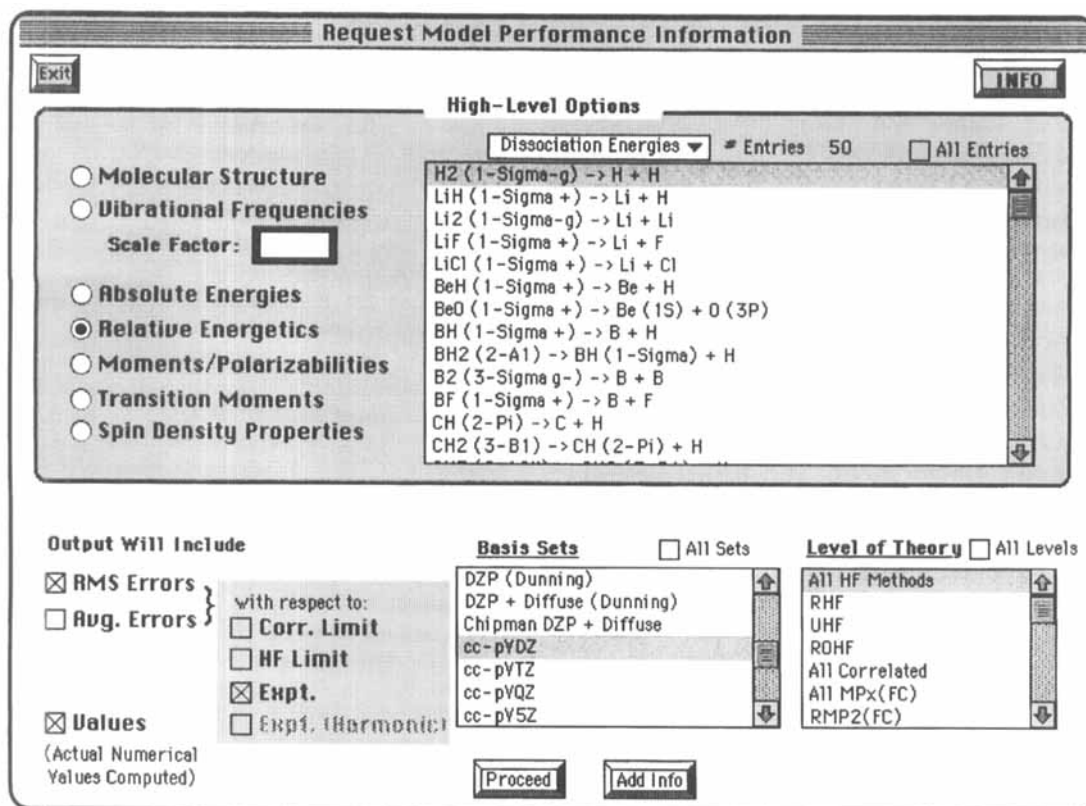


FIGURE 6. The CCIA Methods Performance Database window.

electronic structure tools which incorporate a variety of systematically designed databases tailored for electronic structure methods. The information contained in the databases encompasses applications, hardware platforms, basis sets, and the performance of selected basis set/level of theory combinations in several key areas. As of this writing, the Methods Performance Database contains over 6500 entries, many at higher levels of theory than have previously been published. The Basis Set Database contains 106 sets, as well as a large amount of relevant information on the basis sets.

The databases should ideally function as an integral part of a broader graphical user interface which supports electronic structure applications. The proposed databases complement the existing suite of postprocessing visualization tools which has been available for several years in commercial packages. Prototype basis set, methods performance, and hardware/software databases have been developed.

The tools, which were implemented in the Computational Chemistry Input Assistant prototype,

begin to address the problems arising from the lack of rigorous error bars on *ab initio* methods and the underlying complexities of basis set selection. Although we have chosen to focus on *ab initio* methods, the approach advocated is general and could just as easily be applied to semiempirical or molecular mechanics techniques. By providing this critical information to scientists in the process of setting up calculations, the chances of recovering the desired results from the calculation are enhanced.

Another capability provided by easy access to information on the electronic structure applications is the increased freedom to switch from one program to another, as the need arises, throughout the course of an investigation. While the CCIA prototype began to explore the concept of emphasizing application neutrality in all high-level user interface windows, the UNIX workstation software being developed by the ECCE project extends that concept by abstracting chemical concepts in an object-oriented data model, which also resides in the database.

Acknowledgments

The author thanks Ms. Anne Schur and Ms. Karen Schuchardt of PNL and Dr. Judith Cushing of Evergreen State College for many helpful discussions during the development of CCIA. The members of the Environmental Molecular Sciences Laboratory are thanked for generously donating their time for testing various versions of the software and to Dr. H. Taylor for a careful reading of this manuscript. Help in formatting several of the basis sets came from Mr. Stephen Brunelle and Mr. Stephen Colson. Finally, the author thanks Drs. Thom Dunning, David Woon, and Kirk Peterson for early access to their correlation-consistent basis set results. This research was supported by the U.S. Department of Energy under Contract DE-AC06-76RLO 1830. This work was supported by the Division of Chemical Sciences, Office of Basic Energy Sciences. The Pacific Northwest National Laboratory is operated by Battelle Memorial Institute. We also wish to thank the Scientific Computing Staff, Office of Energy Research, U.S. Department of Energy, for a grant of computer time at the National Energy Research Supercomputer Center. Inquiries about the availability of the ECCE production software which incorporates many of the ideas described in this article should be addressed to Dr. D. R. Jones at the Pacific Northwest Laboratory (dr_jones@?pnl.gov).

References

1. Gaussian 92, Revision A, M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian, Inc. Pittsburgh, PA (1992).
2. GAMESS, M. Dupuis, D. Spangler, and J. J. Wendoloski, University of California (1980). This version is due to M. W. Schmidt, K. K. Baldridge, J. A. Boatz, J. H. Jensen, S. Koseki, M. S. Gordon, K. A. Nguyen, and S. T. Elbert, *QCPE Bull.* **10**, 52 (1990).
3. MOLPRO, H. J. Werner, P. J. Knowles, J. Almlof, R. Amos, S. Elbert, W. Meyer, E. A. Reinsch, R. Pitzer, and A. Stone, Universität Bielefeld, Bielefeld, Germany, University of Sussex, Falmer, Brighton, England (1992).
4. Webster's New World Dictionary, New World Dictionaries/Simon and Schuster, New York, 1982.
5. J. B. Cushing, M. Roa, D. M. DeVaney, and D. Feller, In *Proceedings of the Sixth International Working Conference on Statistical and Scientific Database Management*, Zurich, Switzerland, 1992.
6. J. B. Cushing, *Computational Proxies: An Object Based Infrastructure for Computational Science*, Ph.D. dissertation, Oregon Graduate Institute of Science and Technology, 1994.
7. J. B. Cushing, D. Maier, M. Roa, D. Abel, D. M. DeVaney, and D. Feller, *Proceedings of the Scientific and Statistical Database Workshop*, Charlottesville, VA, 1994.
8. M. W. Feyereisen, G. Fitzgerald, and A. Kormornicki, *Chem. Phys. Lett.*, **208**, 359 (1993).
9. O. Vahtras, J. Almlof, and M. W. Feyereisen, *Chem. Phys. Lett.*, **213**, 514 (1993).
10. SuperCard 2.0, B. Appleton, G. Poppitz, and C. Watson, Allegiant Technologies, Inc. (1995).
11. HONDO, M. Dupuis, A. Marquez, S. Chin, E. Hollauer, S. A. Maluendes, A. Farazdel, S. P. Karna, P. Mougnot, R. L. C. Daniel, K. Dyall, B. Liu, J. D. Watts, H. O. Villar, G. J. B. Hurst, M. Schmidt, M. S. Gordon, S. Elbert, W. Stevens, H. Basch, T. Takada, D. Spangler, J. J. Wendoloski, B. Brooks, W. Laidig, P. Saxe, J. Rys, and H. F. King, IBM Corporation, Dept. MLM/428, Neighborhood Rd., Kingston, NY 12401 (1994).
12. MELDF-X, L. McMurchie, S. Elbert, S. Langhoff, and E. R. Davidson, The University of Washington, Indiana University, and the Molecular Science Research Center (1993). The MELDF suite of programs was substantially modified by D. Feller, R. Cave, D. Rawlings, R. Frey, R. Daasch, L. Nitchie, P. Phillips, K. Iberle, C. Jackels, and E. R. Davidson.
13. SUPERMOLECULE, J. Almlof and M. Feyereisen, The University of Minnesota (1993).
14. NWChem, A Computational Chemistry Package for Parallel Computers, Version 1.1, H.P.C.C. Group, Pacific Northwest National Laboratory, Richland, WA 99352 (1995).
15. MOLCAS, G. Karlstrom, P. A. Malmqvist, B. O. Roos, A. J. Sadlej, and P. O. Widmark, University of Lund, Lund, Sweden (1990).
16. ACES II, J. F. Stanton, J. Gauss, J. D. Watts, W. J. Lauderdale, and R. J. Bartlett, University of Florida (1994).
17. R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kälmeel, *Chem. Phys. Lett.*, **162**, 165 (1989).
18. M. Häser and R. Ahlrichs, *J. Comput. Chem.*, **10**, 104 (1989).
19. M. Häser, R. Ahlrichs, H. P. Baron, P. Weis, and H. Horn, *Theor. Chim. Acta*, **83**, 455 (1992).
20. TX93, P. Pulay, K. Wolinski and coworkers, University of Arkansas (1995).
21. D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.*, **99**, 1914 (1993).
22. D. E. Woon and T. H. Dunning, Jr., *J. Chem. Phys.*, **101**, 8877 (1994).
23. K. A. Peterson, R. A. Kendall, and T. H. Dunning, Jr., *J. Chem. Phys.*, **99**, 1930 (1993).
24. K. A. Peterson, R. A. Kendall, and T. H. Dunning, Jr., *J. Chem. Phys.*, **99**, 9790 (1993).
25. D. J. DeFrees, B. A. Levi, S. K. Polack, W. J. Hehre, J. S. Binkley, and J. A. Pople, *J. Am. Chem. Soc.*, **101**, 4085 (1979).
26. W. J. Hehre, L. Radom, P. v. R. Schleyer, and J. A. Pople, *Ab Initio Molecular Orbital Theory*, Wiley, New York, 1986.
27. D. Feller, The MSRC Ab Initio Benchmark Suite, A PNL Technical Report, PNL-8670 (1993).
28. D. Feller, R. A. Kendall, and M. J. Brightman, The EMSL Ab Initio Methods Benchmark Report, A PNL Technical Report, PNL-10481 (1995).